Human Interaction and Emerging Technologies (IHIET-AI 2025), Vol. 161, 2025, 103–112
https://doi.org/10.54941/ahfe1005901

Wi-Fi Signal Analysis via Smartphones for Estimating Passenger Counts

Mohammed Alatiyyah

Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

ABSTRACT

Smartphones have become integral to daily life, offering innovative applications across various domains. This study introduces a novel method for counting passengers by analyzing Wi-Fi signals emitted by their mobile devices. The research evaluates the effectiveness of leveraging Wi-Fi data to estimate occupancy, addressing a critical issue in public transportation management. The proposed system involves three core processes: signal detection, data filtering, and passenger count estimation. Key results indicate high accuracy in moderately crowded scenarios, with average deviations of 20% from actual counts and accuracy rates between 90% and 100%. However, under high-density conditions, the system tends to overestimate, occasionally doubling the real count. While further research is required to improve precision in such settings, this study lays a foundation for leveraging digital technologies to enhance transportation operations and service delivery.

Keywords: Public transportation, Digital environment, Passenger estimation, Signal capturing, Wi-Fi

INTRODUCTION

Public transportation systems are vital to the functioning of modern cities, serving as a primary mode of travel for millions daily. Consequently, local governments and transportation authorities have allocated substantial resources to enhance their quality and efficiency (Ryu et al., 2020). Simultaneously, the widespread adoption of smartphones, driven by rapid technological progress, has created a highly interconnected digital environment. Once rare among individuals aged 65 and older, smartphones are now widely used across all age groups, including seniors (PewResearchCenter, 2014). These devices have transformed into compact computers with diverse capabilities, such as GPS, compasses, and light sensors. As a result, there is growing interest in leveraging smartphones as research instruments to collect data on individuals' behaviors and activities in various contexts (Farulla et al., 2015; Nicolau et al., 2015; Park et al., 2015; Bergé et al., 2015; Hu et al., 2015). Moreover, the extensive use of digital technologies has fundamentally changed how people interact with their environment. Homes, workplaces, and even vehicles have become "smart," reflecting an increasing dependence on digital tools in everyday life. This evolution has spurred intense competition among corporations, governments, and organizations to adopt cutting-edge technologies and sustain their competitive advantage.

The emergence of open digital environments (ODEs) has recently created promising research opportunities, particularly in the context of public transportation systems such as metros, buses, trams, and stations (Bánhalmi et al., 2021; Hidayat et al., 2020; Wang & Zhang, 2020; Nitti et al., 2015). Despite the widespread availability of real-time schedules and updates for these modes of transport, overcrowding remains a persistent challenge (Wang & Zhang, 2020). The high costs associated with traditional passenger-counting systems on public transportation vehicles have spurred interest in alternative approaches, such as analyzing the digital Wi-Fi environment within these spaces. Specifically, the expenses related to installing and maintaining sensors on metros, buses, or trams are considerable. Thus, a proposed solution involves using more cost-effective technologies to capture and analyze signals emitted by passengers' devices (Nitti et al., 2015).

This study focuses on ODEs, encompassing public spaces such as metro cars, buses, tram stations, shopping centers, and other bustling environments where unexpected data is plentiful. The primary aim is to evaluate whether analyzing the Wi-Fi environment can provide a reasonably accurate estimation of passenger density while minimizing costs. By examining signals emitted by passengers' devices, the study seeks to address key questions about passenger behavior in public transportation contexts, including the extent of Wi-Fi usage, the frequency of smartphone use during transit, and methods to distinguish signals originating from inside versus outside the vehicle. Ultimately, this research seeks to deepen our understanding of the complex interactions within digital environments and their significant influence on everyday life.

Building on the preceding discussion, this study introduces an innovative method for analyzing Wi-Fi signals emitted by smartphones within public transportation environments, such as metro cars, buses, tram cars, and stations, utilizing Wireshark—an advanced network analysis tool. The study explores the use of media access control (MAC) addresses as unique identifiers and the capture of signals from passengers' devices via probe requests. The key contributions of this paper are proposing a cost-effective and scalable approach to estimating passenger counts by leveraging smartphone Wi-Fi signals, developing a signal acquisition and filtering pipeline to enhance accuracy by removing noise and non-mobile device signals, and designing an algorithm that uses received signal strength indication (RSSI) values while accounting for the structural and spatial characteristics of public transport vehicles and stations.

The structure of the paper is as follows. Section 2 provides a comprehensive review of relevant literature. Section 3 details the proposed methodology, Section 4 presents the experimental findings, Section 5 discusses the results, and the conclusions are presented in Section 6.

RELATED WORK

Researchers have explored various methods to address limitations in video surveillance due to privacy concerns and high costs. For example, Zeng et al. (2015) used Wi-Fi Channel State Information (CSI) to analyze shopper behavior, achieving high accuracy, while Maekawa et al. (2014) developed a Bluetooth-based system to detect train congestion with 83% accuracy. Scholz et al. (2015) introduced a Wi-Fi-based fingerprinting system for classifying individuals by height categories, achieving up to 76% accuracy. Other studies, such as Wei et al. (2015) and de Sanctis et al. (2015), explored the use of radio frequency interference and Wi-Fi beacon analysis for human activity recognition, demonstrating innovative methods with varied accuracy levels.

In indoor localization, GPS limitations have led researchers to explore alternative techniques. Lymberopoulos et al. (2015) tested 22 systems and achieved a minimum error of 0.72 m using a 2.4 GHz Phase Offset technique. Chen et al. (2015) proposed BearLoc, a framework integrating sensors and algorithms to simplify development and improve localization accuracy. Sen et al. (2015) developed CUPID2.0, a system achieving 1.8 m accuracy through infrastructure-free implementation, while Meng et al. (2015) introduced semantic translation of Wi-Fi coordinates for identifying store names with over 90% accuracy. Despite efforts like Kocakusak et al.'s (2015) RSSI-based model database, results indicate considerable variability in accuracy across methods, highlighting the need for further refinement in indoor positioning.

Crowd detection methods have also advanced, with researchers moving beyond traditional sensors. Lathia et al. (2014) used smartphones and social media to provide real-time passenger updates, while Zhou et al. (2015) leveraged cellular signals to estimate urban traffic density in Singapore, demonstrating the feasibility of these approaches despite environmental challenges.

Wi-Fi probe requests have emerged as a versatile tool in various applications. Freudiger (2015) analyzed probing frequency in smartphones, while Schaub et al. (2014) developed PriCal, a system using MAC addresses for efficient time management in offices. Studies by Barbera et al. (2013) and Fukuzaki et al. (2014) analyzed large-scale probe data for sociological insights and pedestrian flow trends, respectively. Other works, such as a Wi-Fi-based self-quarantine monitoring system (Guo and Ho, 2022) and multi-story localization using RSSI signals (Magsino et al., 2021), further demonstrate the potential of probe requests for diverse research and practical applications. These studies underscore the promise of Wi-Fi signal analysis for behavioral insights, tracking, and mobility analysis.

METHODOLOGY

This study proposes a system to detect crowding in public transportation by analyzing Wi-Fi signals emitted by passengers' devices. The methodology is tested through an experimental study conducted on Brighton & Hove buses, employing qualitative research techniques to assess its effectiveness. The system comprises three components: signal capture, data filtering, and analysis and estimation, each with specific tasks to ensure accurate data processing.

In the signal capture phase, the system uses Tshark, a streamlined version of Wireshark, to collect Wi-Fi packets from passengers' devices onboard the bus. Key data elements include MAC addresses, device names, SSIDs, and RSSI values, which help identify devices and filter out external or irrelevant signals. Signals are captured live during bus journeys, offering a cost-effective data collection approach compared to methods near bus stops. Filters are applied to eliminate extraneous data such as IP addresses and protocol types, ensuring only relevant information is retained.

The data filtering phase applies several steps to refine the collected data. Devices located outside the bus are excluded based on RSSI values, with only signals stronger than –80 dBm considered valid. Non-mobile devices are also removed, leaving a dataset of signals likely originating from passengers' smartphones. Duplicate MAC addresses are eliminated to ensure the dataset reflects unique devices present on the bus.

In the analysis and estimation phase, an algorithm calculates the frequency of each MAC address appearing in the filtered data. Devices captured in more than 20% of the packets are identified as likely belonging to passengers, and their count serves as an estimate of the number of individuals onboard. The system operates in intervals between bus stops, capturing and processing data iteratively. Experiments reveal that using a 20% threshold provides the most accurate passenger estimates.

This methodology leverages real-time Wi-Fi signal data to provide a scalable, efficient approach to estimating passenger numbers in public transportation, reducing costs and improving accuracy compared to traditional methods.

EXPERIMENTAL RESULTS

The proposed system relies on specific initial values, such as the frequency of signal transmissions from smartphones and signal strength within and outside a bus environment, to ensure functionality and avoid incorrect assumptions. Initial experiments, conducted using a laptop as the capture device in a controlled setting, aimed to detect smartphones on buses, measure WLAN packet transmission frequency, and gather foundational data for system design. Two main experiments were conducted to evaluate the system's performance in passenger counting, with manual counts taken for comparison. These experiments recorded approximately 127,000 unique MAC addresses, reflecting the abundance of digital devices detected. While the system's estimates were often higher than the actual number of passengers, achieving accuracy rates of 90% to 100% in less crowded areas highlighted its potential. The results provide valuable insights into the system's functionality, with detailed findings presented to inform further refinement.

Table 1. Number of MAC address in each bus stop at Experiment 01.

Bus Stop Name	All Data	RSSI Filter	Non-Mobile Filter
Amex Stadium	913	106	59
Brighton University Falmer	734	216	150
Amex Stadium	913	106	4
Falmer Station	116	21	13
Brighton Academy	83	25	10
Coldean Lane	618	196	159
Wild Park	2972	1155	1040
Ringmer Road	39	25	12
Moulsecoomb Way	50	19	10
Bates Estate	125	55	27
Brighton University	181	58	24
Mithras House	125	84	39
Coombe Road	360	79	39
Lewes Road Bus Garage	286	97	60
Melbourne Street	673	268	213
St Pauls Street	1025	719	633
Elm Grove	1585	732	649
Bottom of Elm Grove	499	324	288
De Montfort Road	876	309	261
Bonchurch Road	51	343	296
Queens Park Junction	361	184	138
Baxter Street	124	68	35
The Hanover	131	81	49
Pepper Pot	564	186	148
Albion Hill	258	125	62
Egremont Gate	534	302	232
Park Street	1704	992	858
Gala Bingo Hall	120	49	20
College Place	1000	178	156
County Hospital	155	18	7
Chesham Street	1029	269	184
St Marys Ilall	257	155	135
LiDL Superstore	861	522	486
Roedean Road	1116	600	549
Marina Cinema	421	96	55

In the first experiment, data collection took place on Bus 23 between Amex Stadium and Marina Cinema from 18:38:55 to 19:24:12, capturing over 20,000 unique MAC addresses with varying counts across bus stops. For instance, Wild Park recorded approximately 3,000 unique MAC addresses, although this does not necessarily correspond to the number of passengers, as some signals originated from outside the bus. Table 1 outlines the unique MAC addresses per stop, while Table 2 reveals that 60 MAC addresses were detected three or more times. After applying an RSSI filter to exclude signals from outside the bus, the number of MAC addresses dropped to 8,000, representing a 58% reduction per stop. Further refinement to remove non-mobile devices led to an additional 33% average decrease in MAC addresses per stop, leaving only 41 MAC addresses captured more than twice. Table 3 compares manually recorded passenger counts with system estimates, highlighting high accuracy in some locations and discrepancies in others, which will be addressed in the next section.

Table 2. Number of MAC address in each bus stop at Experiment 01.

Bus Stops	All MAC Address	RSSI Filter	Non-Mobile Filter
35 bus stops (maximum number)	1	1	None
20 to 25	2	2	1
10 to 19	5	5	4
4 to 9	16	16	16
3	33	17	5
2	1204	240	47
Only in a bus stop	19000	7900	6800

Table 3. Results comparison for the number of passengers obtained by the developed system and manual recording in Experiment 01.

Bus Stop Name	Manual	System	Bus Stop Name	Manual	System
Amex Stadium	4	5	De Montfort Road	19	265
Brighton University	5	7	Bonchurch Road	17	298
Amex Stadium	5	4	Queens Park	17	9
Falmer Station	9	6	Raxter Street	16	38
Brighton Academy	9	11	The Hanover	16	51
Coldean Lane	14	5	Pepper Pot	18	150
Wild Park	14	17	Albion Hill	18	5
Ringmer Road	15	13	Egremont Gate	18	234
Moulsecoomb Way	15	11	Park Street	18	9
Bates Estate	15	103	Gala Bingo Hall	21	21
Brighton University	15	24	College Place	23	157
Mithras House	15	53	County Hospital	21	7
Coombe Road	16	41	Chesham Street	19	6
Lewes Road Bus Garage	16	60	St Marys Hall	20	137
Melbourne Street	15	6	LiDL Superstore	18	2
St Pauls Street	19	635	Roedean Road	25	8
Elm Grove	16	650	Marina Cinema	1	1
Bottom of Elm Grove	18	293			

The second experiment was conducted on the Route 7, running from Brighton Marina to Livingstone Road between 09:20:24 and 09:58:52, during which the system captured over 24,000 unique MAC addresses. Table 4 presents the number of MAC addresses recorded at each bus stop, revealing that most were detected at only one stop, while approximately 5% were captured at multiple stops. Table 5 details the MAC addresses captured at multiple bus stops. Applying an RSSI filter reduced the number of MAC addresses by about 80%, with the filtered results shown in Table 4. A subsequent non-mobile filter further decreased the number of MAC addresses by an average of 35% per stop. Finally, Table 6 compares the manually recorded passenger numbers to the system's estimated results, highlighting the system's performance in passenger detection.

Table 4. Number of MAC address in each bus stop at Experiment 02.

Bus Stop Name	All Data	RSSI Filter	Non-Mobile Filter
Brighton Marina	358	47	27
Arundel Road	319	82	57
LiDL Superstore	126	43	19
Sussex Square	264	24	11
St Marys Hall	214	29	16
Chesham Street	130	31	18
County Hospital	4964	342	206
College Place	237	39	27
Gala Bingo Hall	982	133	99
Park Street	275	87	44
Devonshire Place	367	99	58
Law Courts	580	144	108
Old Steine	1002	232	190
Old Steine2	879	201	113
North Street	2812	595	449
Clock Tower	2345	397	281
North Road	2404	449	348
Brighton Station	855	173	107
Compton Avenue	778	221	124
Seven Dials	642	155	97
Osmond Road	272	83	55
Montefiore Road	630	127	49
Lyon Close	639	145	105
Holland Road	420	92	57
Wilbury Villas	593	191	117
Eaton Gardens	571	241	170
Hove Station	663	274	191
Livingstone Road	859	247	166

Table 5. Number of MAC address in each bus stop at Experiment 02.

Bus Stops	All MAC Address	RSSI Filter	Non-Mobile Filter
28 bus stops (maximum number)	1	1	None
20 to 27	5	1	None
10 to 19	7	5	5
5 to 9	37	22	19
3 to 4	98	18	13
2	965	198	89
Only in a bus stop	23000	4200	2800

Table 6. Results comparison for the number of passengers obtained by the developed system and manual recording in Experiment 02.

Bus Stop Name	Manual	System	Bus Stop Name	Manual	System
Brighton Marina	20	5	North Street	27	20
Arundel Road	26	9	Clock Tower	21	12
LiDL Superstore	26	19	North Road	20	20
Sussex Square	28	11	Brighton Station	11	10
St Marys Hall	29	16	Compton Avenue	12	29
Chesham Street	29	18	Seven Dials	18	16
County Hospital	35	18	Osmond Road	20	7
College Place	35	27	Montefiore Road	20	49
Gala Bingo Hall	39	17	Lyon Close	21	105
Park Street	39	7	Holland Road	23	57
Devonshire Place	39	9	Wilbury Villas	20	8
Law Courts	37	13	Eaton Gardens	19	9
Old Steine	31	21	Hove Station	13	5
Old Steine2	31	14	Livingstone Road	13	5

DISCUSSION

This section discusses the findings from Experiments in the context of existing literature, focusing on the factors influencing system performance and identifying opportunities for improvement in Wi-Fi-based passenger estimation systems.

The results from Experiments indicate that the system tends to overestimate passenger numbers in environments with high pedestrian activity or vehicular congestion. This behavior aligns with observations by Mishalani et al. (2016) and Paradeda et al. (2019), who highlighted the impact of external environmental factors on similar systems. These findings underscore the challenges of accurately distinguishing on-bus signals from external ones in crowded urban settings.

Experiment #1 demonstrated how bus stops at traffic signals led to increased capture of MAC addresses from nearby pedestrians, influencing the system's accuracy. This reflects the variability introduced by the bus's movement and external factors, consistent with findings by Nitti et al. (2020) in their iABACUS system. Such scenarios necessitate more robust filtering mechanisms to reduce false positives from nearby devices.

In Experiment #2, the system performed better during peak times, capturing signals more effectively from passengers who had their smartphones' Wi-Fi enabled. However, variability in smartphone usage behavior among passengers—such as devices with Wi-Fi disabled or passengers without smartphones—remained a key challenge, as also noted by Wang and Zhang (2020). These factors highlight the need for systems to account for demographic differences and behavioral patterns in smartphone usage.

The findings from Experiments #1 and #2 contribute to the growing research on Wi-Fi-based passenger estimation systems. While they demonstrate the potential of using Wi-Fi signals for estimating bus occupancy, they also highlight the need for refinement in filtering algorithms and system calibration. Integrating GPS technology and additional data sources, such as multiple signal monitors, could enhance accuracy by minimizing errors from external signals, as suggested by Nitti et al. (2020) and Junior et al. (2022).

To enhance system performance, future improvements should focus on refining filtering algorithms to better differentiate between on-bus and external signals, thereby reducing false positives. Additionally, recalibrating signal capture percentages could help address variations in passenger demographics and smartphone usage behaviors, ensuring more accurate estimations. Integrating GPS technology and multi-monitor setups would provide supplementary contextual data, enabling more precise signal analysis and minimizing errors caused by external sources. These enhancements, informed by experimental findings and existing literature, can significantly improve the reliability and accuracy of Wi-Fi-based passenger estimation systems in urban transit environments.

CONCLUSION

This study introduced a novel system for estimating bus passenger numbers using Wi-Fi signals emitted by smartphones, with a focus on signal detection, data filtering, and passenger count estimation. The system demonstrated high accuracy in moderately crowded conditions, with deviations averaging 20% and accuracy rates of 90% to 100%. However, in high-density environments, it tended to overestimate passenger counts due to external factors such as pedestrian activity and vehicular congestion, as well as variability in smartphone usage behaviors. To address these limitations, future work should refine filtering algorithms to better distinguish on-bus signals from external ones and incorporate GPS technology and multi-monitor setups to enhance contextual accuracy. Adjusting signal capture mechanisms to account for demographic variations and smartphone usage patterns could further improve precision. Expanding the system's application to other high-density scenarios, such as malls or stadiums, and conducting additional testing in diverse operational conditions, will help ensure its scalability and reliability for real-world use.

REFERENCES

Bánhalmi, A., Bilicki, V., Megyeri, I., Majó-Petri, Z., & Csirik, J. (2021). Extracting information from Wi-Fi traffic on public transport. *International Journal of Transport Development and Integration*, 5, 15–27. doi: 10.2495/TDI-V5-N1-15-27.

Bergé, L. P., Dubois, E., & Raynal, M. (2015). Design and evaluation of an 'around the smartphone' technique for 3D manipulations on distant display. *3rd ACM Symposium on Spatial User Interaction (SUI)*, Los Angeles, CA, USA, August 8–9, pp. 69–78. doi: 10.1145/2788940.2788950.

De Sanctis, M., Cianca, E., Di Domenico, S., Provenziani, D., Bianchi, G., & Ruggieri, M. (2015). WiBeCAM: Device-free human activity recognition through Wi-Fi beacon-enabled camera. *2nd Workshop on Physical Analytics*, Florence, Italy, May 22, pp. 7–12. doi: 10.1145/2753497.2753503.

Farulla, G. A., Russo, L. O., Rosa, S., & Indaco, M. (2015). ORIENTOMA: A novel platform for autonomous and safe navigation for blind and visually impaired. *10th International Conference on Design and Technology of Integrated Systems in Nanoscale Era (DTIS)*, Naples, Italy, April 20–24, pp. 1–6. doi: 10.1109/DTIS.2015.7127360.

Hidayat, A., Terabe, S., & Yaginuma, H. (2020). Estimating bus passenger volume based on a Wi-Fi scanner survey. *Transportation Research Interdisciplinary Perspectives*, 6, 100142. doi: 10.1016/j.trip.2020.100142.

Hu, S., Su, L., Liu, H., Wang, H., & Abdelzaher, T. F. (2015). Smartroad: Smartphone-based crowd sensing for traffic regulator detection and identification. *ACM Transactions on Sensor Networks (TOSN)*, 11, 1–27. doi: 10.1145/2806890.

Maekawa, Y., Uchiyama, A., Yamaguchi, H., & Higashino, T. (2014). Car-level congestion and position estimation for railway trips using mobile phones. *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Seattle, Washington, September 13-17, pp. 939–950. doi: 10.1145/2638728.2641553.

Nicolau, H., Montague, K., Guerreiro, T., Rodrigues, A., & Hanson, V. L. (2015). Holibraille: Multipoint vibrotactile feedback on mobile devices. *12th International Web for All Conference (W4A)*, Florence, Italy, May 18-20, pp. 1–4. doi: 10.1145/2745555.2746655.

Nitti, M., Pinna, F., Pintor, L., Pilloni, V., & Barabino, B. (2020). iABACUS: A Wi-Fi-based automatic bus passenger counting system. *Energies*, 13(6), 1446. doi: 10.3390/en13061446.

Park, D. H., Liu, M., Zhai, C., & Wang, H. (2015). Leveraging user reviews to improve accuracy for mobile app retrieval. *38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Santiago, Chile, August 9–13, pp. 533–542. doi: 10.1145/2766462.2767734.

Pew Research Center. (2014). Older Adults and Technology Use. Retrieved from https://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/ [Accessed 10 Oct 2023].

Ryu, S., Park, B. B., & El-Tawab, S. (2020). WiFi sensing system for monitoring public transportation ridership: A case study. KSCE Journal of Civil Engineering, 24, 3092–3104. doi: 10.1007/s12205-020-2120-3.

Scholz, M., Kohout, L., Horne, M., Budde, M., Beigl, M., & Youssef, M. A. (2015). Device-free radio-based low overhead identification of subject classes. *2nd Workshop on Physical Analytics*, Florence, Italy, May 22, pp. 1–6. doi: 10.1145/2753497.2753501.

Wang, Y., & Zhang, X. (2020). Estimating Transit Ridership Using Wi-Fi Signals: An Enhanced Rule-Based Approach. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Toronto, Ontario, Canada, October 11–14, pp. 2425–2431. doi: 10.1109/SMC42975.2020.9282884.

Wei, B., Hu, W., Yang, M., & Chou, C. T. (2015). Radio-based device-free activity recognition with radio frequency interference. *14th International Conference on Information Processing in Sensor Networks*, Seattle, Washington, April 13–16, pp. 154–165. doi: 10.1145/2737095.2737106.

Zeng, Y., Pathak, P. H., & Mohapatra, P. (2015). Analyzing shopper's behavior through Wi-Fi signals. *2nd Workshop on Physical Analytics*, pp. 13–18. doi: 10.1145/2753497.2753517.